International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 11, Issue 06, June 2021)
Manuscript Received: 09 April 2021, Received in Revised form: 20 May 2021, Accepted: 08 June 2021 DOI: 10.46338/ijetae0621_10

Stereo Matching Algorithm Based on Hybrid Convolutional Neural Network and Directional Intensity Difference

Mohd Saad Hamid[1], Nurulfajar Abd Manap[2], Rostam Affendi Hamzah[3], Ahmad Fauzan Kadmin[4]

[1]*Faculty of Electrical and Electronic Engineering Technology, Universiti Teknikal Malaysia Melaka, Malaysia*
[2]*Faculty of Electronic and Computer Engineering, Universiti Teknikal Malaysia Melaka, Malaysia*

***Abstract*—Fundamentally, a stereo matching algorithm produces a disparity map or depth map. This map contains valuable information for many applications, such as range estimation, autonomous vehicle navigation and 3D surface reconstruction. The stereo matching process faces various challenges to get an accurate result for example low texture area, repetitive pattern and discontinuity regions. The proposed algorithm must be robust and viable with all of these challenges and is capable to deliver good accuracy. Hence, this article proposes a new stereo matching algorithm based on a hybrid Convolutional Neural Network (CNN) combined with directional intensity differences at the matching cost stage. The proposed algorithm contains a deep learning-based method and a handcrafted method. Then, the bilateral filter is used to aggregate the matching cost volume while preserving the object edges. The Winner-Take-All (WTA) is utilized at the optimization stage which the WTA normalizes the disparity values. At the last stage, a series of refinement processes will be applied to enhance the final disparity map. A standard benchmarking evaluation system from the Middlebury Stereo dataset is used to measure the algorithm performance. This dataset provides images with the characteristics of low texture area, repetitive pattern and discontinuity regions. The average error produced for all pixel regions is 8.51%, while the nonoccluded region is 5.77%. Based on the experimental results, the proposed algorithm produces good accuracy and robustness against the stereo matching challenges. It is also competitive with other published methods and can be used as a complete algorithm.**

***Keywords*— Convolutional neural network, directional intensity difference, disparity map, matching cost computation, stereo matching algorithm,**

I. Introduction

A disparity map is an output generated by using a stereo matching algorithm. This map is valuable for many applications such as 3D reconstruction, obstacle avoidance, robotics, and navigation. [1] also mentioned the importance of a disparity map and 3D model for the agriculture industry.

The study on disparity estimation has attracted many researchers in the computer vision field[2]. [3] also mentioned the stereo matching algorithm remains a challenging problem in recent years. The stereo matching algorithm generates the disparity map by identifying matching features from the left and right image pair. The disparity values computed in the disparity map represent the difference in pixel locations of matching elements in the left and right images. This map also can help to determine the distance between an object and the stereo camera.

The monocular method is another camera-based method that can also provide depth-related data. A monocular depth method can generate a disparity map using a single image. However, as pointed out by [4], the monocular depth method was less accurate than the stereo-based approach. It is because the multi-view contains a broader amount of information compared to the single-view navigation.

Light Detection and Ranging (LiDAR) is another approach to provide depth-related information. LiDAR uses the laser approach to sense the depth and perform the depth measurement. Even though the LiDAR method can provide accurate 3D points, this method is not very cost-effective and time-consuming [5]. Furthermore, LiDAR also has a constraint on limited spatial density. In terms of computational cost, a vision-based approach will be a more viable option to choose. Due to the highlighted shortcomings in monocular and LiDAR methods, we are inspired to continue working on the stereo vision approach and proposed our stereo matching algorithm.

II. Literature Review

The implementation of the stereo vision algorithm can be divided into two types: local and global approaches. In the local approach, the computation of disparity values is based on pixels' intensity values in a predetermined support region. In contrast, the global method computes the disparity values by taking the full image context.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 11, Issue 06, June 2021)

As a result, it produces a more accurate disparity map but with a higher computational cost.

The crucial steps for the stereo matching algorithm can be illustrated in Figure 1. This formalization of the main steps also mentioned in several other works of literature([6]–[8]). It begins with The stereo vision algorithm begins with matching cost computation. Some researchers utilize the directional intensity difference to compute matching cost [9]. The calculation of absolute difference and the squared difference is also commonly used because of its low computational complexity. This step produced initial matching costs.

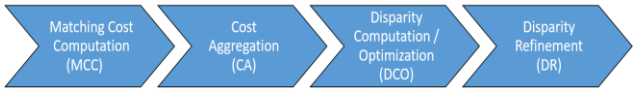

Figure 1 Stereo Matching Algorithm Steps

The second step, cost aggregation, is responsible for decreasing the errors in the initial matching cost. As highlighted in our previous study [8], the edge-preserving filters such as bilateral filter (BF)[10] and guided filter (GF) [11] can achieve the purpose. These filters maintain a good edge while smoothing the input. Thus it provides better results in the aggregation step than the low pass filters (Gaussian and Box filter). This step is commonly associated with the local approach due to its operation over a support region [12].

The third step is in charge of assigning a value to the disparity map. Winner-Take-All (WTA) optimization is the most popular method for this step for the local approach. In WTA, the smallest cost value disparity will be selected for every pixel location [7] to generate an initial disparity map. The third step's initial disparity may still contain errors caused by occlusions, low textures, and invalid matches[13]. So, the final step may contain single or multiple post-processing steps to refine the map. The authors of [11] used the left-to-right consistency check (LRC) process to identify the invalid matched pixels. The median filter is also commonly used for local refinement [14]–[16].

Artificial intelligence (AI) has been a hot subject in the media, with many hypes surrounding it[17]. Machine learning is one of the fascinating areas of artificial intelligence these days. The machine learning algorithm's beauty can tell the computer how to respond or make decisions in specific circumstances without literal instruction code programmed to the computer.

As a subclass of machine learning, deep learning has recently become the driving force behind advancements in the stereo vision field. Deep learning consumes a large amount of data to feed the artificial neural networks for learning purposes. Several researchers highlighted that the implementation of deep learning improved the execution of the task. [18] described that, for recognition tasks, conventional stereo vision could not outperform human results, but the implementation of deep learning will improve their algorithm's performance. [19] also mentioned the deep learning enhances the accuracy for disparity estimation tasks.

We can categorize deep learning implementation on stereo vision into two ways. The first approach combines deep learning with a traditional handcraft algorithm (MC-CNN-acrt [14]). MC-CNN-acrt architecture outperforms other stereo matching conventional methods on Middlebury [20] benchmarking systems. [14] construct and train their convolutional neural network (CNN) based network using binary classification to solve matching cost computation steps. [15] mentioned that CNN used in stereo matching because of its capability to extract features and vigorous radiometric difference.

The second approach is the pure deep learning end-to-end network. This approach does not require any handcrafted algorithm. The end-to-end style deep learning network performs all stereo matching stages in one combination network. GC-Net [21] uses 2D CNN to form cost volumes and used soft argmin layer to regress the disparity values. [22] introduce PSMNet and produced a faster and more accurate disparity map than GC-Net.

A recent hybrid method between learning and handcraft algorithm also published by [23] as illustrated in Figure 2. The authors remodelled deep learning network based on PSMNet[22] into their deep learning network,PSMNU[23]. PSMNU produces a disparity map and predictions of aleatoric uncertainties.

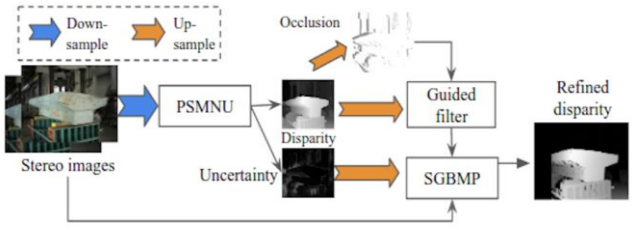

Figure 2 Overview of SGBMP Algorithm[23]

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 11, Issue 06, June 2021)

The initial disparity output is then refined using a Guided filter. The authors also propose a handcrafted method to enhance their disparity map using SGBMP [23]. It is based on the Semi Global Block Matching method to produce their final disparity map. Their final Middlebury evaluation results were also discussed with other published methods and compared in this paper.

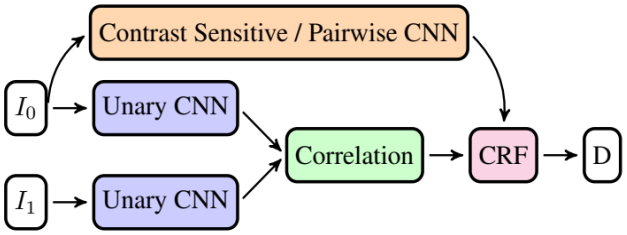

Figure 3 Hybrid CNN-CRF Architecture [24]

A hybrid network between the CNN and learning-based conditional random field (CRF) model was introduced [24], as illustrated in Figure 3. The authors transformed the handcraft CRF into a learning model and combined it with CNN to form an end-to-end learning network to produce a disparity map. Their *Unary-CNN* extract features from image pair (left image(I_0) and right image(I_1)). Both features extracted from I_0 and I_1 will be compared using the *Correlation* layer to generate matching cost volume. The *Pairwise-CNN* will estimate contrast-sensitive pairwise costs. The learned-based CRF used both cost volumes (unary and pairwise) to create a disparity map. Their results on Middlebury datasets also discussed in this paper and labelled as JMR(based on Middlebury online benchmarking system).

As mentioned by [25], the end-to-end method still has some demerits in the ill-posed region and computationally expensive. Additionally, [14] also pointed out that the raw disparity map generated by a convolutional neural network (CNN) prone to errors in the low texture and the occluded region. Several authors ([3], [11], [15], [26]) also highlighted the problem with radiometric changes.

This may affect the accuracy output produced by the stereo matching algorithm. So, this paper will present our proposed stereo matching algorithm. This paper's main contribution is the hybrid CNN fused with directional intensity information for the matching cost computation stage to combat the mentioned causes of errors.

III. The Proposed Method

We present our proposed algorithm as categorized in [6]. Summary view of our algorithm illustrated in Figure 4. Our proposed algorithm contains several techniques grouped into four stages as follows:

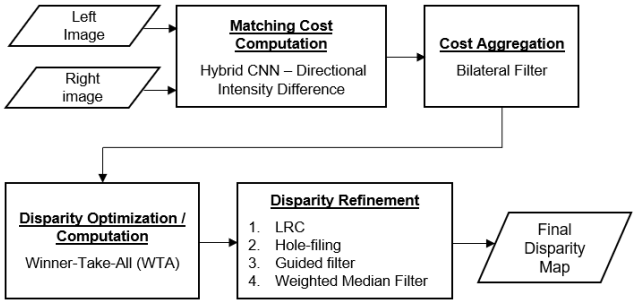

Figure 4 The Proposed Stereo Vision Algorithm Steps

A. Matching Cost Computation

The CNN model implemented was inspired by the MC-CNN-acrt architecture. Our current Siamese-based CNN model performs better in this paper than our implemented CNN model [27]. We improved the matching cost computation step in this paper by tuning the CNN hyperparameters. We also fused the improved CNN model with directional intensity information to enhance the accuracy of matching costs. We changed the model into seven layers model. Finally, we improve the classification part of our CNN model into a more comprehensive architecture. The proposed CNN part for this stage is illustrated in the following Figure 5.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 11, Issue 06, June 2021)

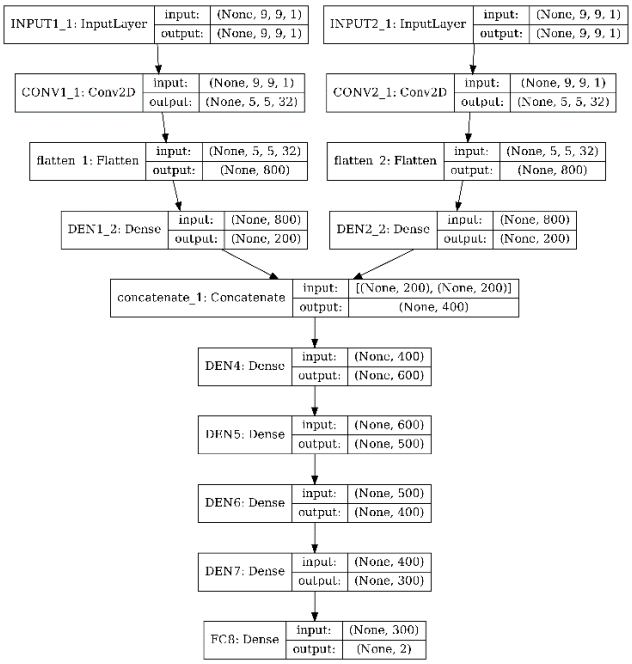

Figure 5 Proposed CNN Architecture

Layer 1 of the model contains a convolutional layer where the input of the layer comes from the input layer that produces 9x9 input patches from the left and right image pairs. The convolutional layer in layer 1 of the model's left and right arm served as feature extractors. The layer's output will be flattened into a 1-dimensional feature vector before supplied to layer 2. Layer 2 contains a fully connected (FC) layer or dense layer that produces 200 neurons. Similar to MC-CNN-acrt architecture mentioned in [8], the feature vectors from left and right are concatenated and supplied to the classification part. The next layer in the classification part, DEN4, is a fully connected layer or dense layer. It contains 600 neurons while neurons of the following layers DEN5, DEN6, and DEN7 are set to 500, 400, and 300 neurons. The final layer FC8 remains unchanged, set to 2 neurons. This CNN model is broader and shorter as compared to our previous model [27].

The network in Figure 5 will provide a binary classification of good or bad matching. Based on Equation (1), C_{CNN} reflects the cost value for all disparities d at each pixel position p.

$$C_{CNN}(p,d) = -s\left(P_N^L(p), P_N^R(p-d)\right) \quad (1)$$

Input patches of NxN size from left and right image, P_N^L and P_N^R were supplied to the CNN as illustrated in Figure 5. As mentioned earlier, the patch size used is 9x9.

Another improvement done in this matching cost computation stage is the directional intensity difference information. The following Equation (2) contains the directional intensity difference part.

$$\begin{aligned} C_{DI}(p,d) \quad &= \alpha \cdot min(\| I_L(p) - I_R(p) \|, \tau_1) \\ &+ (1-\alpha) \cdot min(\| \nabla_x I_L(p) - \nabla_x I_R(p) \|, \tau_2) \end{aligned} \quad (2)$$

The $I(p)$ is the color vector of a pixel at position p. While the ∇_x is the directional intensity difference in the x-direction. α balances the directional intensity difference values, while τ_1 and τ_2 are truncation values. This paper improves the matching cost step by combining CNN-based cost volume with the directional intensity difference-based cost volume, $C_{DI,}$ as in the following Equation (3).

$$C_{IM}(p,d) = C_{DI}(p,d) + \lambda \cdot C_{CNN}(p,d) \quad (3)$$

The parameter λ will balance the C_{DI} cost volumes. For the parameter, the optimum value chosen for λ is 0.034. The matching cost volume cost from this stage defined as C_{IM}.

B. Cost Aggregation

In the second stage, we refine the raw matching costs to create a more accurate disparity map. It is because the initial cost volume generated from the previous stage is prone to noises. We aggregate cost volumes in this stage using a Bilateral filter (BF). The BF is responsible for performing smoothing operations while keeping the edges sharp on the matching costs. The BF will aggregate the raw matching cost volume and defined in the following equations (4) and (5) [10].

$$BF[I]p = \frac{1}{W_p} \sum_{q \epsilon S} G_{\sigma_s}(\| p - q \|) G_{\sigma_r}(| I_p - I_q |) \quad (4)$$

$$W_p = \sum_{q \epsilon S} G_{\sigma_s}(\| p - q \|) G_{\sigma_r}(| I_p - I_q |) \quad (5)$$

Where Wp is the normalization factor. I_p and I_q represent the intensity at pixel position p and q, respectively. p represents the (x, y) coordinates pixel of interest while the neighboring pixel coordinate q is within support region S.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 11, Issue 06, June 2021)

The two main parameter, σ_s act as the spatial parameter of the kernel, and σ_r responsible for the range parameter. The higher value of σ_s will smooth larger feature while the σ_r is the second term to include the effect color difference around p. Increasing the value of σ_r too large will turn the filter closer to the Gaussian filter. However, according to [10], no smoothing effect will occur if any parameter value is set to 0. The aggregated cost at the end of this step denoted as CAGG in the following Equation (6).

$$C_{AGG}(p,d) = BF[I]C_{IM}(p,d) \quad (6)$$

C. Disparity Computation / Optimization

Based on our previous literature [8], we employ a simple yet effective Winner-Take-All (WTA) optimization to compute the disparity map in this stage. As per the following Equation (7), WTA optimization will produce an initial disparity map, $d(p)$.

$$d(p) = \underset{d \epsilon d_r}{\operatorname{argmin}}\left(C_{AGG}(p,d)\right) \quad (7)$$

The sets of disparity value of d_r obtained from the ground truth. The $d(p)$ is the chosen disparity value for position $d(p)$, which represents 2D coordinates (x, y).

D. Disparity Refinement

This stage consists of several steps. Firstly, we employ LRC to identify invalid pixels. Then the detected invalid pixel is replaced with a valid pixel using the hole-filling process. The hole-fileing process described as pseudocode in Figure 6.

```
InitD is the initial disparity map after LRC check
xDP contains all invalid disparity position (x,y - coordinate) from LRC check
For every y-coordinate in xDP
        Slice disparity values along x-axis into L_arm and R_arm array based on InitD
        Lptr is pointer for disparity value in L_arm
        Set Lptr to point rightmost disparity value in L_arm
        Rptr is pointer for R_arm
        Set Rptr to leftmost disparity value in R_arm
        Scan disparity value in L_arm from right to left to find nearest valid value
        Scan disparity value in R_arm from left to right to find nearest valid value
        IF Lptr point to invalid disparity value
            InitD[x,y] set to disparity value pointed by Rptr
        ELSE IF Rptr point to invalid disparity value
            InitD[x,y] set to disparity value pointed by Lptr
        ELSE
            InitD[x,y] set to the smallest disparity value between Lptr and Rptr
END For
Return InitD with new disparity value
```

Figure 6 Pseudocode for Hole-filing

After that, we implemented a guided image filter (GIF) due to its good edge-preserving capabilities. The filter kernel for GIF implemented in this paper defined in Equation (8).

$$G_{p,q}(I) = \frac{1}{|w|^2} \sum_{q \epsilon W_k} \left(1 + \frac{(I_p - \mu_k)(I_q - \mu_k)}{\sigma_k^2 + \varepsilon}\right) \quad (8)$$

where I the guidance image and p represent the (x; y) coordinates pixel of interest. Another pixel position, q, denote the neighbouring pixel in the support region wk. Then, σ and μ are the variance and mean of the intensity values, respectively. The parameter, ε is used as a control element for the smoothness term. Then we refined the disparity map using GIF defined as d_{GF} in Equation (9).

$$d_{GF}(p) = G_{p,q}(I)d(p) \quad (9)$$

The Weighted Median (WM) filter was implemented to remove any existing outliers in the disparity map. We use the following cosine similarity weight function as defined in (10) for the weighted median filtering.

$$W_p^{cos} = \frac{I_p \cdot I_q}{\| I_p \| \cdot \| I_q \|} \quad (10)$$

$$d_f(p) = W_p^{cos} h(p) d_{GF}(p) \quad (11)$$

After the WM filter implemented, the final disparity map generated from the whole proposed algorithm represented by $d_f(p)$ in Equation (11). The next section will provide an overview of the proposed method's performance quantitatively and qualitatively.

IV. Results And Discussion

We performed several experiments using the training dataset of the Middlebury v3 online benchmarking system [20]. The dataset contains indoor-type stereo image pairs with very accurate ground truth as reference. The experiment conducted using a personal computer (PC) platform to examine the proposed method's performance. We used the Keras library with Tensorflow as the back-end using Python programming for the deep learning network training. We performed the training for 500 epochs with a batch size of 128, which will save the best performing weightage for the network upon completion. The learning rate is set to 0.0001 using Adam optimizer.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 11, Issue 06, June 2021)

The training images from Middlebury datasets are split into two parts for network training and validation. The OpenCV[28] library is also used in the proposed algorithm for image acquisition and post-processing operations.

To show the competitiveness of the proposed algorithm, we also performed a comparison with other published methods, MC-CNN-acrt[14], PSMNet_ROB[22], JMR[24], SGBMP[23], MC-CNN-WS[29], HGIF[30], and LS_ELAS[31]. The quantitative comparison performed as illustrated in Table 1 and 2. We also conducted a qualitative comparison with the methods as in Figure 7. The metrics used in the comparison is the average error percentage for invalid disparity values in the whole image and nonoccluded region (denoted as *All* and *NonOcc* error, respectively).

Table 1 shows the comparison of *All* errors between the methods mentioned above. We are comparing methods that utilize the Siamese-based network (MC-CNN-acrt [14] and MC-CNN-WS[29]) and our approach. We can see that our method outperformed both methods. Our results on *All* error is 8.51%, while for MC-CNN-acrt and MC-CNN-WS are 11.80% and 13.70%, respectively. This show significant improvement our proposed method as compared to other Siamese based implementation.

Comparing our result to other hybrid algorithms such as JMR[24] and SGBMP[23]. The output produced by the proposed algorithm also outperformed the other two hybrid algorithms. The JMR algorithm had an average error of 9.57%, and SGBMP made an 11.20% error based on quantitative data on TABLE 1.

TABLE 1
RESULTS FROM MIDDLEBURY BENCHMARK – *ALL* ERROR

Methods	Images															
	Adirondack	ArtL	Jadeplant	Motorcycle	MotorcycleE	Piano	PianoL	Pipes	Playroom	Playtable	PlaytableP	Recycle	Shelves	Teddy	Vintage	Weight Avg.
Proposed	3.82	7.57	35.5	6.48	6.47	5.65	5.58	10.2	10.7	4.2	4.07	3.61	9.99	3.04	9.63	8.51
JMR	2.17	18.00	24.70	5.98	6.90	6.14	7.27	11.00	17.50	8.18	7.44	2.96	7.81	8.98	10.30	9.57
SGBMP	6.50	9.33	56.80	5.04	5.43	4.77	14.80	7.85	7.62	10.60	3.78	3.19	5.00	3.35	30.00	11.20
MC-CNN-acrt	4.24	18.70	34.10	7.21	7.22	6.00	9.35	13.50	18.30	9.71	9.37	4.64	6.62	9.31	21.60	11.80
LS_ELAS	9.31	5.90	64.50	7.24	7.65	6.25	9.69	12.80	10.10	23.90	4.27	7.39	8.48	2.98	14.00	12.90
PSMNet_ROB	8.83	13.90	68.40	8.26	9.16	5.89	10.50	14.40	9.38	5.54	5.52	4.98	11.60	3.87	9.66	13.30
MC-CNN-WS	5.73	20.50	36.30	9.39	9.37	8.13	16.10	16.70	18.70	11.50	10.10	5.05	9.83	11.00	20.80	13.70
HGIF	7.96	21.17	33.00	12.62	12.92	18.11	33.09	21.49	23.79	21.23	14.28	10.33	32.64	11.19	30.94	18.71

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 11, Issue 06, June 2021)

TABLE 2
RESULTS FROM MIDDLEBURY BENCHMARK – *NONOCC* ERROR

Methods	Images															
	Adirondack	ArtL	Jadeplant	Motorcycle	MotorcycleE	Piano	PianoL	Pipes	Playroom	Playtable	PlaytableP	Recycle	Shelves	Teddy	Vintage	Weight Avg.
JMR	0.92	2.18	6.01	1.26	1.27	2.21	4.03	2.12	1.94	2.20	1.65	1.30	5.51	1.15	3.73	2.30
MC-CNN-acrt	0.76	2.49	16.30	1.27	1.27	1.83	5.07	2.29	2.27	3.11	3.03	2.48	4.41	1.07	14.80	3.81
MC-CNN-WS	1.66	4.27	12.80	2.26	2.18	3.21	11.70	4.27	3.49	3.78	3.31	1.83	7.02	2.00	14.30	4.63
Proposed	2.76	6.43	17.5	4.35	4.19	4.88	5.17	6.79	5.78	3.58	3.47	3.05	9.4	2.59	8.33	5.77
SGBMP	3.87	4.96	29.30	3.45	3.89	3.82	14.40	3.94	5.09	9.74	2.70	2.91	4.64	1.80	26.10	7.25
PSMNet_ROB	7.32	9.69	44.50	5.55	6.12	5.01	9.82	9.86	7.33	4.40	4.43	3.73	11.10	3.44	8.07	9.60
LS_ELAS	8.46	3.83	41.10	5.12	5.80	5.54	8.97	7.44	8.76	22.40	3.47	6.93	8.26	2.29	13.10	9.66
HGIF	5.73	10.85	19.67	8.45	8.47	14.02	29.32	9.85	15.18	16.71	10.98	8.28	31.63	5.57	26.97	12.94

The average error on the nonoccluded region, *NonOcc*, is illustrated in **TABLE *2***. The overall results obtained show that the proposed algorithm performed competitively among the other algorithms. The proposed algorithms ranked fourth in **TABLE *2***, where the proposed algorithm produced output with 5.77% of error. The proposed algorithm performed better than the hybrid method SGBMP algorithm (7.25%) and the end-to-end network-based algorithm, PSMNet_ROB(9.60%).

When we compared the PianoL image (image with different lighting conditions) to the original Piano image, we found that the proposed method performed better than the other algorithms when there are radiometric changes. The amount of error percentage difference between PianoL and Piano image generated by the proposed method is the smallest among all other algorithms. This result valid for the result as illustrated in **TABLE *1*** and **TABLE *2***.

The comparison shows that the proposed method is robust against radiometric changes and performed better than other algorithms.

In terms of qualitative comparisons, we compared the output sample from the Playtable image generated by the proposed algorithm with the output from other algorithms, as illustrated in Figure 7. Overall comparison, the disparity map produced by the proposed algorithm exhibit a smooth output compared to other algorithms except for the bottom-left of the output. However, this error also exists in other algorithms in the figure. We also focused on the low texture area at the bottom of the table in image pairs.

Based on our observation, the output produced by the proposed algorithms smoother as compared to the output produced by the SGBMP algorithm. The qualitative comparison reflects quantitative results recorded in **TABLE *1***. The other sets of disparity maps generated based on the Middlebury dataset are also displayed in Figure *8*.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 11, Issue 06, June 2021)

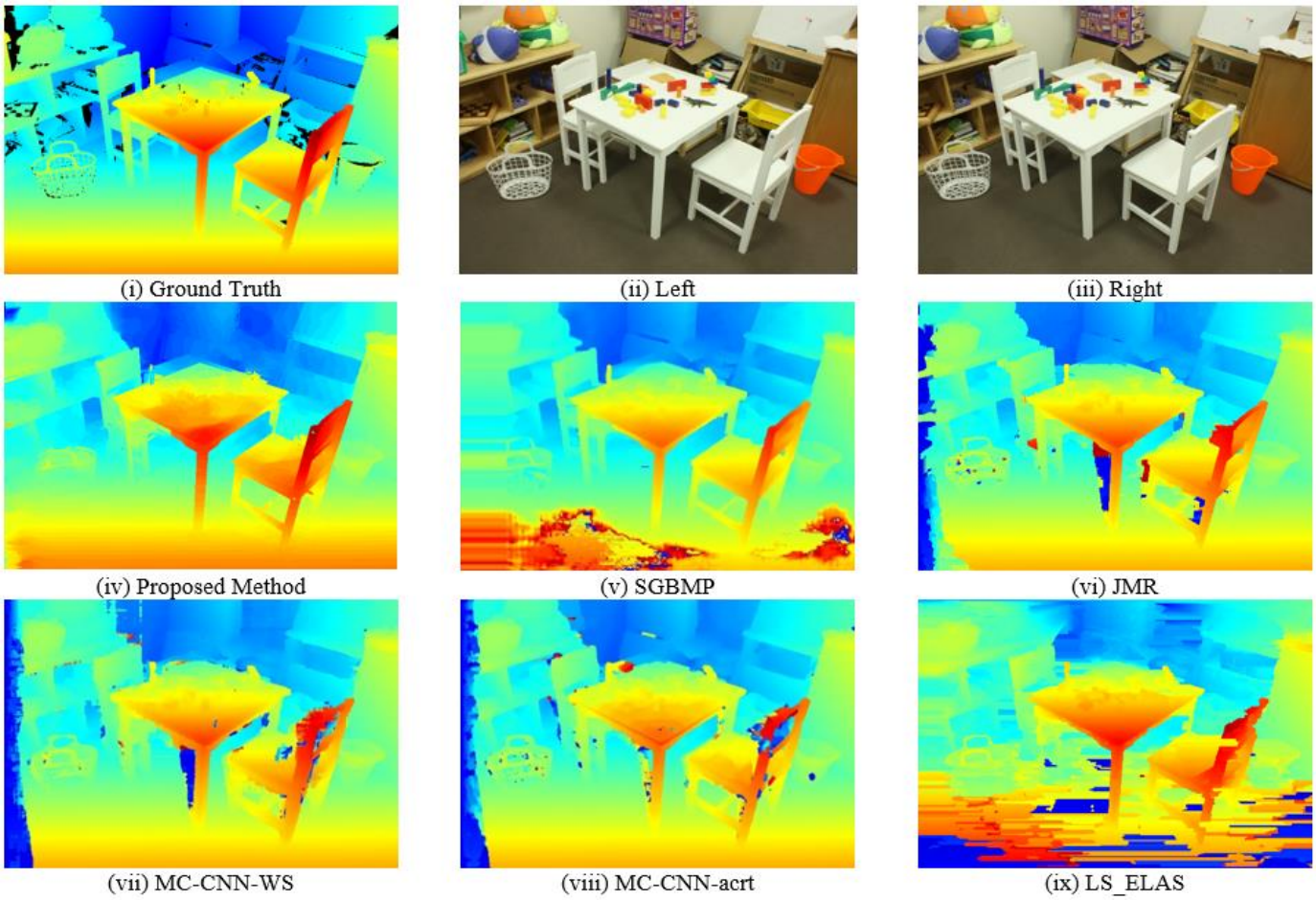

Figure 7 Middlebury – Playtable Image Comparison

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 11, Issue 06, June 2021)

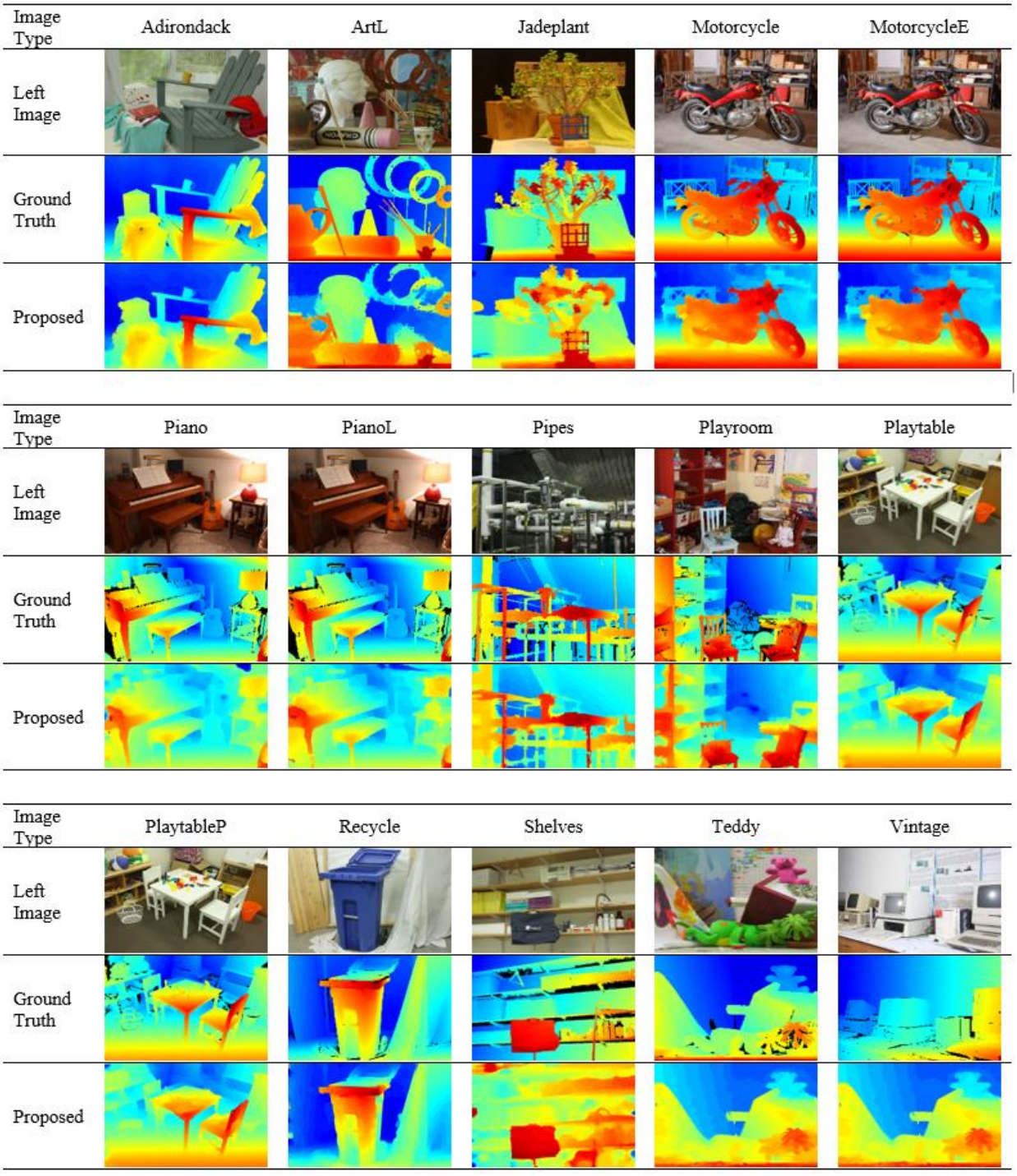

Figure 8 The Generated Disparity Map using Proposed Algorithm

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 11, Issue 06, June 2021)

V. CONCLUSION

In conclusion, we demonstrated our proposed stereo matching algorithm, which contains a hybrid learning-based method combined with a handcrafted technique to perform matching cost computation. We found that the initial cost volume created in the first stage still contains noises. The cost aggregation step was implemented using BF to reduce noises and errors. BF is proven effective in our experiments as an edge-preserving filter to maintain the sharp edge while smoothing the input. To compute the initial disparity map, we used simple yet effective WTA optimization to produce the initial disparity map. Several post-processing steps in the final stage, including LRC, the hole filing process, GIF, and WMF, delivered the final disparity map. Based on the final disparity map sample in Figure 7, the proposed algorithm can improve the disparity map's accuracy in the low texture region while maintaining the object edge. Based on the overall performance, as shown in Table 1 and Table 2, the proposed algorithm can perform competitively compared to other published methods based on the Middlebury stereo benchmarking system. The Piano and PianoL image results in the tables proved that the proposed algorithm is robust against radiometric changes compared to other published algorithms. So, for our future work, we will focus on improving and further testing the algorithm on outdoor datasets such as KITTI dataset due to this promising result.

Acknowledgements

This work is supported by the Ministry of Higher Education (MOHE), Malaysia, Universiti Teknikal Malaysia Melaka (UTeM) and sponsored by grant number: (FRGS/1/2020/FTKEE-CACT/F00451).

REFERENCES

[1] A. J. Malekabadi, M. Khojastehpour, and B. Emadi, "Comparison of block-based stereo and semi-global algorithm and effects of pre-processing and imaging parameters on tree disparity map," Sci. Hortic. (Amsterdam)., vol. 247, no. May 2018, pp. 264–274, 2019, doi: 10.1016/j.scienta.2018.12.033.

[2] A. Seki and M. Pollefeys, "SGM-Nets: Semi-global matching with neural networks," in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017, pp. 6640–6649, doi: 10.1109/CVPR.2017.703.

[3] Z. Wang, S. Zhu, Y. Li, and Z. Cui, "Convolutional neural network based deep conditional random fields for stereo matching," J. Vis. Commun. Image Represent., vol. 40, no. Part B, pp. 739–750, 2016, doi: 10.1016/j.jvcir.2016.08.022.

[4] N. Smolyanskiy, A. Kamenev, and S. Birchfield, "On the importance of stereo for accurate depth estimation: An efficient semi-supervised deep neural network approach," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., vol. 2018-June, pp. 1120–1128, 2018, doi: 10.1109/CVPRW.2018.00147.

[5] G. Yang, J. Manela, M. Happold, and D. Ramanan, "Hierarchical Deep Stereo Matching on High-Resolution Images," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 5510–5519, doi: 10.1109/CVPR.2019.00566.

[6] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo," Int. J. Comput. Vis., vol. 47, no. 1, pp. 7–42, 2002, doi: 10.1109/SMBV.2001.988771.

[7] R. A. Hamzah and H. Ibrahim, "Literature survey on stereo vision disparity map algorithms," J. Sensors, vol. 2016, 2016, doi: 10.1155/2016/8742920.

[8] M. S. Hamid, N. A. Manap, R. A. Hamzah, and A. F. Kadmin, "Stereo matching algorithm based on deep learning: A survey," J. King Saud Univ. - Comput. Inf. Sci., 2020, doi: 10.1016/j.jksuci.2020.08.011.

[9] K. Zhang, Y. Fang, D. Min, L. Sun, S. Yang, and S. Yan, "Cross-Scale Cost Aggregation for Stereo Matching," IEEE Trans. Circuits Syst. Video Technol., vol. 27, no. 5, pp. 965–976, 2017, doi: 10.1109/TCSVT.2015.2513663.

[10] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand, "Bilateral filtering: Theory and applications," Found. Trends Comput. Graph. Vis., vol. 4, no. 1, pp. 1–73, 2009, doi: 10.1561/0600000020.

[11] S. Zhu, Z. Wang, X. Zhang, and Y. Li, "Edge-preserving guided filtering based cost aggregation for stereo matching," J. Vis. Commun. Image Represent., vol. 39, pp. 107–119, 2016, doi: 10.1016/j.jvcir.2016.05.012.

[12] Y. Xu, Y. Zhao, and M. Ji, "Local stereo matching with adaptive shape support window based cost aggregation," Appl. Opt., vol. 53, no. 29, p. 6885, 2014, doi: 10.1364/ao.53.006885.

[13] P. Brandao, E. Mazomenos, and D. Stoyanov, "Widening siamese architectures for stereo matching," Pattern Recognit. Lett., vol. 120, pp. 75–81, Apr. 2019, doi: 10.1016/j.patrec.2018.12.002.

[14] J. ˇZbontar and Y. LeCun, "Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches," J. Mach. Learn. Res., vol. 17, pp. 1–32, 2016, doi: 10.1186/s13568-015-0106-7.

[15] S. Wen, "Convolutional neural network and adaptive guided image filter based stereo matching," in 2017 IEEE International Conference on Imaging Systems and Techniques (IST), 2017, no. 1, pp. 1–6, doi: 10.1109/IST.2017.8261530.

[16] H. Xu, "Stereo matching and depth map collection algorithm based on deep learning," in IST 2017 - IEEE International Conference on Imaging Systems and Techniques, Proceedings, 2018, vol. 2018-Janua, no. 1, pp. 1–6, doi: 10.1109/IST.2017.8261504.

[17] F. Chollet, Deep Learning with Python, 1st ed. Greenwich, CT, USA: Manning Publications Co., 2017.

[18] D. Ciresan, U. Meier, J. Schmidhuber, D. Cires, and U. Meier, "Multi-column Deep Neural Networks for Image Classification," in 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, no. February, pp. 3642–3649, doi: 10.1109/CVPR.2012.6248110.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 11, Issue 06, June 2021)

[19] X. Yang, L. He, Y. Zhao, H. Sang, Z. L. Yang, and X. J. Cheng, "Multi-Attention Network for Stereo Matching," IEEE Access, vol. 8, pp. 113371–113382, 2020, doi: 10.1109/ACCESS.2020.3003375.

[20] D. Scharstein et al., "High-resolution stereo datasets with subpixel-accurate ground truth," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2014, vol. 8753, pp. 31–42, doi: 10.1007/978-3-319-11752-2_3.

[21] A. Kendall et al., "End-to-End Learning of Geometry and Context for Deep Stereo Regression," in Proceedings of the IEEE International Conference on Computer Vision, 2017, vol. 2017-Octob, pp. 66–75, doi: 10.1109/ICCV.2017.17.

[22] J.-R. Chang and Y.-S. Chen, "Pyramid Stereo Matching Network," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, doi: 10.1109/CVPR.2017.730.

[23] Y. Hu, W. Zhen, and S. Scherer, "Deep-Learning Assisted High-Resolution Binocular Stereo Depth Reconstruction," in 2020 IEEE International Conference on Robotics and Automation (ICRA), 2020, pp. 8637–8643, doi: 10.1109/ICRA40945.2020.9196655.

[24] P. Knöbelreiter, C. Reinbacher, A. Shekhovtsov, and T. Pock, "End-to-end training of hybrid CNN-CRF models for stereo," Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 1456–1465, 2017, doi: 10.1109/CVPR.2017.159.

[25] X. Song, X. Zhao, H. Hu, and L. Fang, "EdgeStereo: A Context Integrated Residual Pyramid Network for Stereo Matching," Lect. Notes Comput. Sci., vol. 11365, pp. 20–35, 2019, doi: 10.1007/978-3-030-20873-8_2.

[26] R. A. Hamzah, A. F. Kadmin, M. S. Hamid, S. F. A. Ghani, and H. Ibrahim, "Improvement of stereo matching algorithm for 3D surface reconstruction," Signal Process. Image Commun., vol. 65, 2018, doi: 10.1016/j.image.2018.04.001.

[27] M. S. Hamid, N. A. Manap, R. A. Hamzah, and A. F. Kadmin, "Converged classification network for matching cost computation," Int. J. Adv. Sci. Technol., vol. 29, no. 6 Special Issue, 2020.

[28] G. Bradski, "The OpenCV Library," Dr. Dobb's J. Softw. Tools, 2000.

[29] S. Tulyakov, A. Ivanov, and F. Fleuret, "Weakly Supervised Learning of Deep Metrics for Stereo Reconstruction," in Proceedings of the IEEE International Conference on Computer Vision, 2017, vol. 2017-Octob, pp. 1348–1357, doi: 10.1109/ICCV.2017.150.

[30] C. Zhu and Y.-Z. Chang, "Hierarchical Guided-Image-Filtering for Efficient Stereo Matching," Appl. Sci., vol. 9, no. 15, 2019, doi: 10.3390/app9153122.

[31] R. A. Jellal, M. Lange, B. Wassermann, A. Schilling, and A. Zell, "LS-ELAS: Line segment based efficient large scale stereo matching," in 2017 IEEE International Conference on Robotics and Automation (ICRA), 2017, pp. 146–152, doi: 10.1109/ICRA.2017.7989019.